Early Childhood Longitudinal Study, Birth Cohort: Combined Kindergarten— 1st Grade Field Test (ECLS-B, K-1) Field Test Report Memo #4

Child Assessment

Michelle Najarian
Kyle Snow

Conducted by:
RTI International
P.O. Box 12194
Research Triangle Park, NC 27709

Early Childhood Longitudinal Study, Birth Cohort: Combined Kindergarten— 1st Grade Field Test (ECLS-B, K-1) Field Test Report Memo #4

Child Assessment

March 2006

Project No. 08116.013

Michelle Najarian
Kyle Snow

Conducted by:

RTI International
P.O. Box 12194
Research Triangle Park, NC 27709

Please see comments from the Assessment memos, as much of the text here is the same.

Contents

Tables

Figures

1.   Objective of the Field Test Assessments

The Early Childhood Longitudinal Study, Birth Cohort (ECLS-B) kindergarten and first grade (K-1) field test was conducted between fall 2005 and winter 2006. The field test was smaller in scale than the field test for the preschool round and was focused on a small number of specific instruments and procedures. These have been described in previous field test reports. This report describes field test results relevant to the development of the child direct assessment for use in the kindergarten rounds of the National Study.[1] A more limited field test was possible because data on procedures and many of the child assessment items could be drawn from the preschool national study. Additionally, the child assessment to be used in the kindergarten rounds of the National Study drew items from the Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K) instruments that have known properties and, therefore, need only a limited amount of testing. The child assessment goals of the K-1 field test were to develop a lowest level for the kindergarten wave child cognitive assessment and to include items from the ECLS-K kindergarten assessment to enable a direct linkage between the ECLS-K and ECLS-B dataset as the ECLS-B study children enter kindergarten. The first goal was driven by the fact that the children in the ECLS-B will be somewhat younger than the children in the ECLS-K[2] and therefore the ECLS-B might have a greater percentage of children who are functioning below the average range. The second goal was to identify items for inclusion in the final child assessment that allowed for linking to both the preschool data from the ECLS-B cohort and the kindergarten data from the ECLS-K, permitting both longitudinal and cross-cohort comparisons.

One of the objectives of the ECLS-B was to assess the birth cohort longitudinally as well as to compare the birth cohort to the kindergarten cohort and, thus, be able to study changes in children born 8 years apart. Consequently, for the national study, it is necessary to establish two assessment scores for each child: a vertical score and a horizontal score. The vertical (i.e., longitudinal) scale measures change in the birth cohort over time and thus links the ECLS-B preschool and kindergarten scores. The horizontal (i.e., cross-sectional) scale allows for a comparison of the performance of the birth cohort in kindergarten to that of their peers in the kindergarten cohort. The field test sample served to link the assessment scores from the national preschool test with scores from the ECLS-K kindergarten-first grade battery. For field test children whose ages matched those in the ECLS-K sample in kindergarten and first grade (i.e., mostly those with January to September births), we would expect to see percentage correct on the ECLS-K items comparable to the ECLS-K fall kindergarten and fall first grade national

---

[1] At the time of the field test, the intention was to conduct data collection during both the kindergarten and first grade year. The data collection plan has since changed to focus on collecting data from children in fall 2006, and then again in fall 2007 for children entering kindergarten for the first time. However, the intent of the field test was to inform data collection for the proposed K-1 data collection rounds.

[2] Not all of the ECLS-B study children will be enrolled in kindergarten during the first kindergarten round (Fall '06) due to age requirements and other factors. Thus, this group as a whole will be, on average, younger than the children from the base year of the ECLS-K, since all the ECLS-K children were age eligible for kindergarten.

statistics. To the extent that the percentage correct differs from the previous statistics and from the preliminary national preschool round results, we would adjust the statistics to estimate a smooth transition from the preschool- to the kindergarten-level items.

In order to ensure the linking of the ECLS-B with the ECLS-K, constructs assessed by the ECLS-B preschool cognitive assessment were mapped onto constructs assessed by the ECLS-K K-1 cognitive assessment. This mapping was important because the two assessments are not identical, and the map will make it possible for researchers to chart children's development across cohorts in areas such as phonological sensitivity, print awareness, and pattern understanding. It is also expected that the crosswalk between the two assessments will help in the development of the scores.

It should be noted that the link between data from the ECLS-B Kindergarten '06 data collection and the ECLS-K may only be possible for the subset of children who are in kindergarten, not those who are still in preschool. This is because there may be a floor effect using the ECLS-K scale for those children who are in preschool during the kindergarten round. It should be possible to make comparisons to the ECLS-K once these children are in kindergarten, during the ECLS-B Kindergarten '07 data collection.

1.1 Field Test Instrument

The ECLS-B K-1 field test child assessment instrument was designed to be consistent with the structure of the final instrument to be used in ECLS-B Kindergarten '06 and Kindergarten '07 data collection rounds, which is structurally similar to the assessment used in the ECLS-K K-1 rounds. It includes a routing test and two or more second-stage tests that are appropriate for the individual child's developmental level. Separate tests are provided for language/literacy and mathematics skills. The selection of the second-stage form depends on each child's performance on the routing test. For purposes of the field test, children were tested only in English. A parallel Spanish language form will be developed for the national data collection, and children will be tracked to it based upon their performance on a small number of English competency items and an indication that Spanish is spoken in the home.

One of the goals of the ECLS-B program is to link the performance of children in the ECLS-B cohort with that of children in the ECLS-K cohort when both are in their kindergarten year. Using the same instruments will facilitate this comparison. However, the children seen in ECLS-B will be somewhat younger than those seen in ECLS-K during kindergarten year data collection[3]. The ECLS-K cohort was recruited from kindergarten programs and thus included some children who would have already turned 6 at the time of the data collection. However it is expected that approximately 25 percent of the ECLS-B cohort will not be enrolled in kindergarten at the time of 2006 data collection. In addition, because twin births and low birth

---

[3] Since ECLS-B study children were sampled from births in 2001 whereas study children in the ECLS-K study were sampled from kindergarten classrooms.

weight children were oversampled in ECLS-B, there is likely to be a greater percentage of children who are functioning below average for their age and/or have a higher prevalence of special health needs. As a result, there is a need for more items for children at the lower end of the ability spectrum. Moreover, because the ECLS-K collected data at the end of first grade, in addition to the beginning of first grade, there are many items on the ECLS-K K-1 assessment that are expected to be too difficult for the ECLS-B cohort. Thus, the differences in the age distribution of the cohorts and the timing of the data collection require us to supplement ECLS-K items with items that are at a lower level. To account for these differences in cohorts and data collection timelines, the planned design for the child assessment in ECLS-B Kindergarten ('06 and '07) includes the use of forms one and two from the ECLS-K assessment and the creation of a new bottom level using items from the ECLS-B preschool assessment. The field test instrument was designed to (1) select and test ECLS-B preschool items that are candidates for inclusion in the Kindergarten assessment and (2) examine the performance of children on a subset of the ECLS-K items so that the resulting child assessment scores will be on a scale that is interpretable.

Because the ECLS-K K-1 child assessment does not have a level easy enough for the youngest children in the ECLS-B kindergarten rounds, it was necessary to develop a lower level for the cognitive assessment drawn from items that were a part of the ECLS-B preschool field test. Item statistics (i.e., item difficulties and item response theory [IRT] parameters) obtained from the preschool field test guided our decisions regarding item selection for the new lowest level. Items at the upper end of the difficulty range were considered for inclusion, particularly those that were appropriate for the oldest children in the field test sample. In addition to considering the item statistics, we chose items from each of the constructs identified in the cognitive assessment framework for the study.

1.2   Field Test Sample

The K-1 Field Test sample came from two sources: the existing field test sample and a small newly recruited sample. The existing field test sample for ECLS-B consisted of 1,247 families who participated in the preschool field test. Roughly half of these had been in the field test sample since the 9-month field test; the other half was recruited specifically for the preschool field test. From these, we selected 229 cases for the K-1 Field Test. These cases were referred to as the "K-1 Field Test longitudinal sample." Table 1 shows the types of children selected from this sample in order that children likely to be at the lower end of the range of functioning (i.e., the very low birth weight [VLBW] children and those born in the last months of 2000) were represented to the same extent as those at a higher level of functioning (i.e., those born earlier in 2000 and/or enrolled in kindergarten).

Table 1.   Revised selection strata for the kindergarten cohort of the K-1 field test

Strata	Dates of birth	Number selected	Home visits completed
Very low birth weight (VLBW) children from the original field test longitudinal sample	January to March 2000	16	15
Non-VLBW children from the original field test longitudinal sample or preschool supplemental sample	January to March 2000	104	79
Very low birth weight (VLBW) children from the original field test longitudinal sample	April 2000	2	2
Non-VLBW children from the original field test longitudinal sample or preschool supplemental sample	April 2000	11	11
Children from the preschool supplemental sample	April to June 2000	15	13
Children from the preschool supplemental sample	July to September 2000	38	33
Children from the preschool supplemental sample	October to December 2000	43	38
Newly recruited sample of kindergartners	1999	—	39

— Not available, see text for details.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), 9-month National Data Collection, Preschool Field Test, and Combined Kindergarten-1st Grade Field Test.

In selecting the cases from the existing ECLS-B Field Test sample, eight of the 17 field test Primary Sampling Units (PSUs) were chosen in which to work (Denver, CO; Fredericksburg/Reston/DC; St. Paul, MN; Pittsburgh, PA; Philadelphia, PA; Greenville, SC; North Austin, TX; and Lumberton, NC). These were selected to represent a variety of regions of the country, both urban and rural areas, and different types of school districts.

This sample was supplemented with a newly recruited sample of kindergartners who were born in 1999. These children were older than those in the existing field test sample and enhanced our ability to assess the school recruitment and teacher data collection methodology. Also, this sample was needed to mirror the national sample in 2007, in which there will be a group of children who are older but enrolled in kindergarten for the first time (due to delayed kindergarten entry or a late birthday). Newspaper advertisements were run in the same eight PSUs from which the longitudinal sample was drawn, asking interested parents of kindergarteners to call a toll-free number to enroll in the study. Callers completed a brief screening questionnaire to collect the date of birth of the child and to confirm that the child was enrolled in kindergarten and that both the parent and child spoke English. These procedures resulted in a supplemental sample of 41 kindergarteners, of which 39 completed home visits and the child assessment.

First-grade cases born in 1999 were also newly recruited for the field test. The recruitment of this sample happened concurrently with recruitment of the supplemental kindergarten sample and followed the same protocol.

The original sampling plan called for a smaller sample from the longitudinal sample of children with January to March 2000 birthdates. This number was increased during the field test period in order to include additional older children in the sample, due to difficulties in recruiting the supplemental sample of cases. The proportion of first-graders was reduced because the first grade data collection was dropped in favor of a second kindergarten wave, to be fielded in fall 2007.

The final field test sample of 344 children came from two sources: existing field test sample ($N = 229$) and a small newly recruited sample ($N = 115$). The longitudinal sample was supplemented with children born in 1999 who were attending kindergarten ($N = 41$) or first grade ($N = 74$). By the end of the field test, 303 children (88 percent) had completed the child assessment, and contributed data for the item analyses. Figure 1 presents a graphic of the combined kindergarten and first grade samples.

Figure 1. Combined kindergarten and first-grade samples
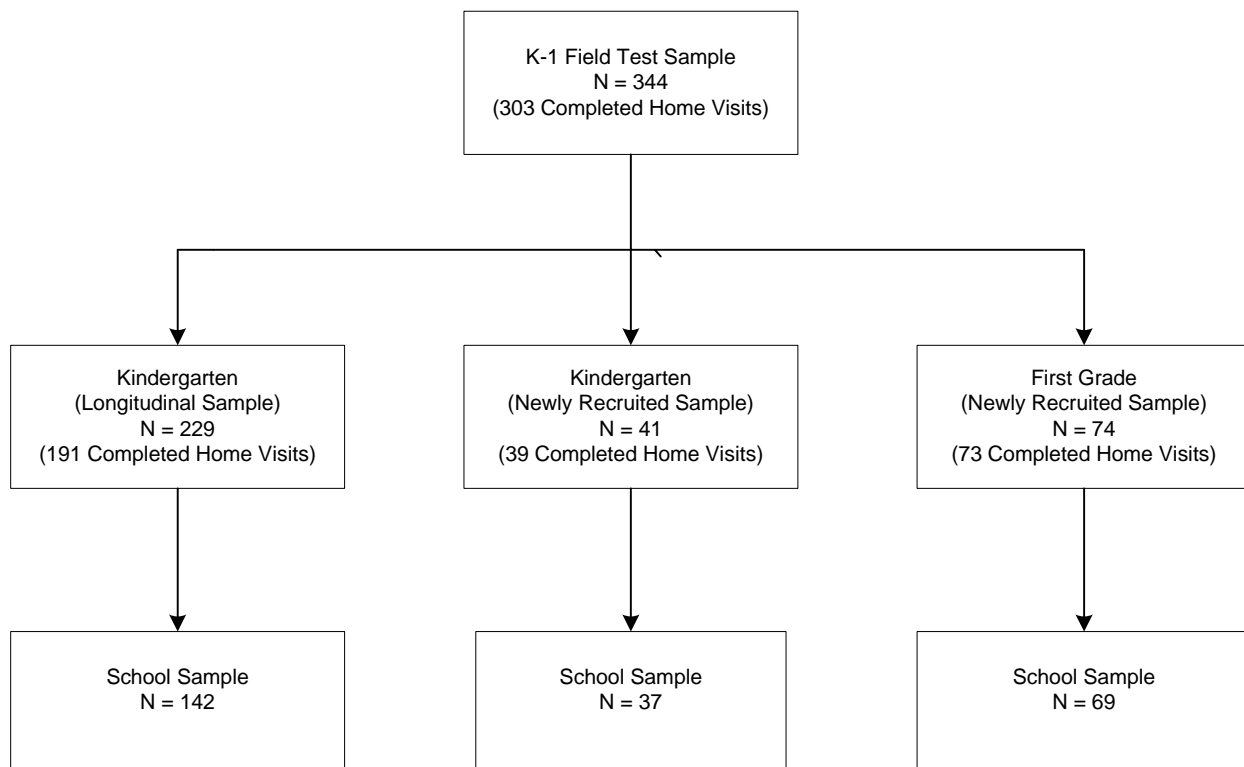
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), 9-month National Data Collection, Preschool Field Test, and Combined Kindergarten-1st Grade Field Test.

1.3 Analytic Methods

An evaluation of item quality was conducted by Educational Testing Services (ETS) using item response theory (IRT) and differential item functioning (DIF). The first approach

provides information on item difficulty and the relationship of individual items to the construct as a whole; the second analyzes differential performance by subgroup.

Unlike the more traditional, classical item analysis approach, which considers percentage correct (P+) for each item as a central indicator of item quality, IRT analysis extends the analysis by simultaneously taking into account omits and the possibility of guessing. The IRT difficulty parameter ("b") for each item is on the same scale as the ability estimate (theta) for each child, allowing for matching a set of test items to the range of ability of sampled children. The IRT ("a") parameter "discrimination" is analogous to the r-biserial of classical item analysis showing the relationship between each item and the total.

DIF analysis shows whether there is any relative advantage or disadvantage of a test item for children of different population subgroups who are matched on overall performance. Items are classified as "A" (no significant difference) to "C" (substantial and statistically significant difference for the focal group compared with a reference group). In general, "C" items are automatically deleted from an item pool.

All of the items considered for inclusion in the kindergarten assessment have been used before, either in the ECLS-B preschool or ECLS-K K-1 large-scale operational tests, so item quality concerns have already been addressed to some extent. At this stage, particular attention was paid to item-quality issues with the data from the national preschool sample, since for most of these items, previous analysis had been done only on the relatively small field test sample. Although not all of the national preschool data were available for inclusion in the K-1 field test, a sizeable percentage was available and thus provided a better look at item analysis statistics than was possible previously. This is especially true for DIF analysis, because low Ns for some racial/ethnic groups precluded DIF analysis for some groups at the field test stage. For example, two items were found to exhibit C-level DIF. "Point to the longest line" (preschool reference MR3 M-I-GE-41B)) and "Point to the bicycle that is between the other bicycles" (preschool reference MR20 (M-I-GE-37B)) exhibited DIF against blacks and Asians, respectively, and were therefore not included in the design of the kindergarten national assessment. (These items did not show DIF in the small preschool field test sample.)

Item quality and potential for use in the kindergarten test forms was performed by reviewing all of the available information for each item. For example, an item with $P+ = .25$ (a quarter of children answering correctly) and IRT $b = 2.0$ would appear to be a difficult item, potentially suitable for a difficult-level test form. But a low r-biserial (below about .30 or .35) or a relatively flat IRT "a" parameter (below .50 or so) would suggest a weak relationship between the item and the test as a whole. In other words, while the item is difficult, it is not useful in differentiating different levels of cognitive skills.

In order to measure children's cognitive skills accurately, it is important that each child receive a set of test items that is appropriate to his or her skill level. The selection of items for the kindergarten form brings together two sets of information: the difficulty parameters for each

of the items in the pool, and the range of ability expected in the kindergarten round. Calibration of these two pieces of information *on the same scale*, so that they may be used in conjunction with each other, was accomplished by means of IRT analysis. (In-depth discussions of the application of IRT to longitudinal studies may be found in the ECLS-K psychometric reports.)

IRT calibration was carried out by pooling the following datasets together:

- ECLS-K fall kindergarten (approximately 18,000 cases)
- ECLS-K spring kindergarten (approximately 19,000 cases)
- ECLS-K fall first grade (data collected only for a subsample of about 5,000)
- ECLS-K spring first grade (approximately 16,000 cases)
- ECLS-B preschool round (7050 cases available at the time of analysis)
- ECLS-B K-1 field test (303 cases)

The overlapping items shared by two or more datasets serve as anchors, so that parameters for items and samples from different assessments are all on a common scale. Data from the preschool national data collection provides item quality statistics for the preschool items, on a larger sample than was possible in the preschool field test, including DIF comparisons for groups whose field test Ns were too small. The preschool data also provides item difficulty parameters, on a scale calibrated jointly with the K-1 items. Data from the K-1 field test supplies a link between the preschool items and the ECLS-K kindergarten/first grade assessment[4], enabling them to be put on a common scale which is necessary for measurement of longitudinal gains and cross-assessment comparisons. The large samples from the ECLS-K kindergarten and first grade national data collections also serve to stabilize parameter estimates that would be unreliable if only the small sample of the fall 2005 K-1 field test sample were available.

2. Language and Literacy Items

2.1 Completion Rates

All children ($N = 303$) completed the language and literacy items on the K-1 field test. An important caveat to this rate, however, is that stop rules were used so that children were routed out of certain sections of the assessment early if they were unable to correctly answer the first five items presented in the series. In practice, this resulted in children completing items that estimated their skill level, while those items above their performance level were not presented. Therefore, children did not complete items that were "too hard" beyond established criteria within the assessment. Such rules result in keeping the child engaged by ending the assessment

---

[4] The ECLS-B K-1 field test links the preschool National Study with the ECLS-K K-1 because the ECLS-B K-1 field test contains items from both assessments.

once the items became more difficult, before the child would likely become distressed by an inability to complete these harder items.

2.2 Completion Time

The median completion time for the language and literacy assessment was 11.5 minutes. The mean, median, and inter-quartile ranges for completion time is given in table 2.

Table 2. Child assessment length

Child assessment cases—language and literacy			
N	Mean time (in minutes)	Median time (in minutes)	Inter-quartile range
303	11.92	11.5	12.18

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

2.3 Literacy Item Analysis

We began by using IRT ability estimates to define target difficulty ranges for different forms of the test. The ability (theta) estimates for the ECLS-K fall kindergarten sample, from the pooled analysis above, is used as a best estimate of the ECLS-B fall kindergarten sample, with a mean = –0.654, and standard deviation = 0.630. To the extent that the pooled ECLS-K fall kindergarten data is representative, three standard deviations above and below the mean (about -2.54 to +1.24) should include about 99 percent of the ECLS-B kindergarten sample. This range also covers the (pooled) preschool ability estimate to two standard deviations (0.568) below its mean (–1.35), which provides coverage to those children in the lower ability range, possibly such as those not entering kindergarten in fall 2006.

The range of -2.54 to +1.24 defines not only the ability range of the children, but also the corresponding difficulty (b) parameters of the items required for the assessment. Of course, the parameter estimates are not precise or permanent, and will change to some extent in the national administration. There are many reasons for possible changes in the parameter estimates: the assortment of items and the order in which they will be given, the number and location of practice items, discontinue rules, etc. As a precaution against encountering floor and ceiling effects in the national assessment (sets of items that are *all* too hard or too easy for some of the children taking them) we recommend extending the difficulty range of the items at both the low and high ends. We have included some items with (b) parameters below –2.54 on the proposed low second stage form, and some above +1.24 on the high form. Overall, a total of 88 items across the routing, level 0, level 1, and level 2 forms were selected for inclusion in the assessment (because some items are common across forms, the number of individual items is 74).

We reviewed the psychometric characteristics of the items and deleted any items that were unsatisfactory with respect to the quality criteria described above. Then we sorted the items according to category and presentation. For example "Basic Skills" items were presented in

several different formats. We reviewed the difficulty statistics for the items within each content/presentation type, and classified each set as suitable for the routing test, the low, middle or high form, according to whether the majority of the items in each set had an appropriate spread of difficulty. We compared different presentations of the same content, and where there was redundancy, selected the item sets with the strongest characteristics. In general we ordered the items so that content was presented in increasing order of average difficulty (although most content areas had a spread of difficulty within them) as we grouped item types.

The form design was based on what was used in the ECLS-K K-1 assessments, with several modifications. The ECLS-B kindergarten assessment does not require the very difficult items included on the high form of ECLS-K, which had to challenge children who were in the spring of first grade. The ECLS-K high form was therefore removed, and the ECLS-K middle and low forms became the ECLS-B Level 2 and Level 1 forms, respectively. Again, several modifications were made, such as rearranging, adding, and removing items. A Level 0 form was assembled, which includes several items from the ECLS-B preschool national assessment, in addition to some ECLS-K low form items. The routing test was also modified to include all items from the ECLS-K proficiency levels 1 to 5, in addition to four easier PPVT items.

The distributions of thetas described above define the range of ability to be targeted by the test forms. The IRT difficulty parameters for the pool of available items are calibrated on the same scale as the abilities. Thus the process of choosing test items relies on matching the difficulty of the items to the ability of the test takers. To optimize the measurement accuracy of the tests, the selected items should be approximately equally spaced along the ability/difficulty scale. Table 3 below shows the peak difficulty ranges for the set of items to be administered to each child, that is, the routing test plus one second stage form (50 percent of the abilities expected to fall in the Routing + Level 1 category, 25 percent lower, and 25 percent higher), the number of items in the peak range, the full range of difficulty per form, and the total number of items per form. As indicated above, to avoid floor and ceiling effects, the item difficulties in each of the forms extend beyond the peak difficulty ranges. (The items outside the peak range are a result of including the full range of routing items, the intentional addition of items to extend difficulties beyond the peak range, and the addition of items to provide overlap between forms needed to support development of a common score scale.)

Table 3.   Peak and full difficulty ranges, routing + second stage

	Routing + Level 0	Routing + Level 1	Routing + Level 2
Total number of items at each level, including the 24 routing items	40	45	51
Peak difficulty range	-2.55 to -1.07	-1.07 to -0.23	-0.23 to +1.24
Number of items in peak range	15	22	28
	7 Routing	6 Routing	10 Routing
	8 Level 0	16 Level 1	18 Level 2
Full range of difficulty	-3.04 to +1.25	-2.25 to +1.25	-2.25 to +1.64

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

2.3.1   Recommendation for the National Assessment

The child assessment in language and literacy in the Kindergarten '06 and '07 rounds will again use a two-stage design, with all children being administered the routing form, which will direct them to the appropriate second-stage form based upon their ability level. The recommended child language and literacy assessment includes 24 routing items, as well as 16 items at level 0, 21 items at level 1, and 27 items at level 2. These items are drawn based upon their psychometric strength, as revealed through IRT analyses. The language and literacy items for each form of the assessment, with parameters for discrimination, difficulty, and guessing, are shown in table 4. Note that in Table 4, the first 5 items on form level 0 constitute part of the language screen that is administered prior to the other language and literacy items at the beginning of the direct child assessment; however, child responses to these items will be incorporated into child scores for level 0.

Several factors contributed to the total number of items recommended for inclusion in the national assessment. As described above, the initial target was to match the number of total items for the Levels 1 and 2 forms to be consistent with the ECLS-K Low and Middle forms. First, all of the ECLS-K proficiency items were incorporated into the routing form. For the remaining number of items, the item difficulties, frameworks, psychometric characteristics and number of linking items were all taken into consideration to design the assessment forms. There are 6 fewer items on the Level 1 form as on the Level 2, and this reduction in number of items was kept (nearly) consistent between the Levels 1 and 0 forms, with 5 fewer items on the Level 0 as on the Level 2.

Table 4. Item run parameters for language and literacy items to be included in ECLS-B Kindergarten '06 and '07 child assessment, by routing form and item order

Proposed form	Short name	item description	Combined run parameters		
			Discrimination	Difficulty	Guessing
ROUTING					
R-1	painting	PV: Point to painting	0.954	-2.254	0.102
R-2	Yawning	Point to yawning	1.224	-1.655	0.263
R-3	dressing	PV: Point to dressing	0.877	-1.652	0.107
R-4	fruit	PV: Point to fruit	1.073	-1.490	0.291
R-5	LETREC D	What letter is this? (D)	1.982	-1.183	0.000
R-6	LETREC M	What letter is this? (m)	1.951	-1.133	0.000
R-7	LETREC F	What letter is this? (F)	2.231	-1.122	0.000
R-8	LETREC T	What letter is this? (t)	2.105	-1.023	0.000
R-9	BEG P	Beginning sound: Paste	1.398	-0.582	0.000
R-10	BEG R	Beginning sound: Run	1.866	-0.548	0.000
R-11	BEG L	Beginning sound: Liquid	1.831	-0.512	0.000
R-12	END L	Ending sound: Mill	1.744	-0.279	0.000
R-13	END F	Ending sound: Shelf	1.456	-0.233	0.000
R-14	BEG B	Beginning sound: Barnacle	1.156	-0.133	0.000
R-15	END P	Ending sound: Keep	1.323	-0.058	0.000
R-16	END D	Ending sound: Glad	1.364	0.104	0.000
R-17	RUNS	Read aloud: runs	2.761	0.470	0.000
R-18	WENT	Read aloud: went	2.722	0.592	0.000
R-19	DOWN	Read aloud: down	3.337	0.646	0.000
R-20	JEEP	Read aloud: jeep	2.681	0.655	0.000
R-21	BACKPACK	books in backpack	2.835	1.087	0.186
R-22	LISTEN	listen to story	4.277	1.151	0.155
R-23	RIDEBIKE	ride her bicycle	3.920	1.239	0.229
R-24	SIZES	shapes and sizes	4.126	1.252	0.139
LEVEL 0					
0-1	OpenHand	Simon says open your hand	1.398	-2.762	0.000
0-2	TouchEar	Simon says touch your ear	1.589	-2.540	0.000
0-3	LiftFoot	Simon says lift one foot	1.765	-2.429	0.000
0-4	Book	What is this? (Book)	1.243	-3.036	0.000
0-5	BookWhat	What can you do with it? (Book)	1.907	-2.240	0.000
0-6	Cup	What is this? (Cup)	1.150	-2.728	0.000
0-7	CupWhat	What can you do with it? (Cup)	1.603	-2.235	0.000
0-8	BEGIN	Begin story: go which way?	0.799	-1.196	0.000
0-9	NEXTLINE	End of line: go where?	0.950	-0.752	0.000
0-10	STORYEND	Where does story end?	1.039	-0.804	0.000
0-11	BEG BIKE	Point to first letter of bike	1.949	-1.083	0.240
0-12	S2	What sound does the letter S make?	1.585	-1.255	0.000
0-13	Nn	What letter makes the Nnn sound? (N)	2.103	-0.928	0.345
0-14	Shoe	Point to shoe	0.871	-2.088	0.093
0-15	Playgrnd	Say playground without ground	1.642	-0.784	0.000
0-16	MOM	Read aloud: mom	1.858	-0.281	0.000

See notes at end of table.

Table 4. **Item run parameters for language and literacy items to be included in ECLS-B Kindergarten '06 and '07 child assessment, by routing form and item order—Continued**

| Proposed form	Short name	item description	Combined run parameters		
			Discrimination	Difficulty	Guessing
LEVEL 1					
1-1	BEGIN	Begin story: go which way?	0.799	-1.196	0.000
1-2	NEXTLINE	End of line: go where?	0.950	-0.752	0.000
1-3	STORYEND	Where does story end?	1.039	-0.804	0.000
1-4	BEG WORD	Point to where "caught" begins	0.795	-0.015	0.000
1-5	QUESMARK	What is this (question mark)?	0.916	-0.038	0.000
1-6	BEG BIKE	Point to first letter of bike	1.949	-1.083	0.240
1-7	Nn	What letter makes the Nnn sound? (N)	2.103	-0.928	0.345
1-8	truck	Which of these pictures starts with the same sound as truck (trip)	1.096	-0.819	0.392
1-9	sneeze	Which picture starts with the sn sound like snail? (sneeze)	1.857	-0.769	0.196
1-10	sock	Which of these pictures starts with the same sound as sock (sun)	1.785	-0.636	0.384
1-11	Run	Which word begins with the same sound as run (rope)	1.502	-0.552	0.310
1-12	Airport	Say airport without air	1.578	-0.524	0.000
1-13	Heat	Say heat without t	1.004	-0.246	0.000
1-14	KAYLAFLY	What happened to Kayla?	0.531	-0.850	0.000
1-15	COULDNOT	What in story could not happen?	0.718	-0.923	0.000
1-16	COULD	What in story could happen?	0.479	-0.321	0.000
1-17	AWARDING	PV: Point to awarding	0.781	-0.404	0.265
1-18	TIME	PV: Point to time	1.157	-0.341	0.176
1-19	JOGGING	PV: Point to jogging	0.991	-0.308	0.137
1-20	MOM	Read aloud: mom	1.858	-0.281	0.000
1-21	YELLOW	Read aloud: yellow	1.526	-0.153	0.000
LEVEL 2					
2-1	KAYLAFLY	What happened to Kayla?	0.531	-0.850	0.000
2-2	COULDNOT	What in story could not happen?	0.718	-0.923	0.000
2-3	COULD	What in story could happen?	0.479	-0.321	0.000
2-4	AWARDING	PV: Point to awarding	0.781	-0.404	0.265
2-5	TIME	PV: Point to time	1.157	-0.341	0.176
2-6	JOGGING	PV: Point to jogging	0.991	-0.308	0.137
2-7	MOM	Read aloud: mom	1.858	-0.281	0.000
2-8	YELLOW	Read aloud: yellow	1.526	-0.153	0.000
2-9	YOU	Read aloud: you	2.189	0.133	0.000
2-10	FISHING	Read aloud: fishing	4.100	0.714	0.000
2-11	THROUGH	Read aloud: through	2.935	1.401	0.000
2-12	CATCH	Read aloud: catch	3.078	0.749	0.000
2-13	TOIL	Read aloud: toil	1.946	1.640	0.000
2-14	mike	Say mike without /k/	0.928	0.214	0.000
2-15	tiger	Say tiger without /g/	0.970	0.857	0.000
2-16	winter	Say winter without /t/	1.204	0.872	0.000
2-17	OR PIG	OR: If "dig" started with "p"	1.691	0.274	0.000
2-18	OR SAT	OR: If "mat" started with "s"	2.218	0.311	0.000

See notes at end of table.

Table 4. Item run parameters for language and literacy items to be included in ECLS-B Kindergarten '06 and '07 child assessment, by routing form and item order—Continued

Proposed form	Short name	item description	Combined run parameters		
			Discrimination	Difficulty	Guessing
2-19	OR TAIL	OR: If "sail" started with "t"	2.498	0.427	0.000
2-20	OR HAND	OR: If "band" started with "h"	2.556	0.512	0.000
2-21	BOYBIRD	Point to boy on a bike	3.137	0.664	0.190
2-22	CANINBAG	Point to can going into bag	1.806	0.735	0.230
2-23	KITN BED	Point to kittens on the bed	2.722	0.740	0.173
2-24	GIRLREAD	Point to girl reading	2.598	1.012	0.251
2-25	KIM CAT	What is Kim?	3.843	0.957	0.504
2-26	NEEDHOME	What does Kim need?	4.687	0.583	0.170
2-27	LIKE DRY	Why not outside on rainy days?	3.369	1.117	0.250

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

Because each child will receive items in the routing form in addition to the appropriate second-stage form, we would expect item difficulty across the possible combinations (routing + form 0, routing + form 1, routing + form 2) to demonstrate greater difficulty in the higher forms than the lower forms. Item difficulty for the routing test items plus items in level 0 (see figure 2), routing plus level 1 (see figure 3), and routing plus level 2 (see figure 4) are shown below.

Figure 2. Item difficulties for language and literacy items in the routing and level 0 child assessment forms
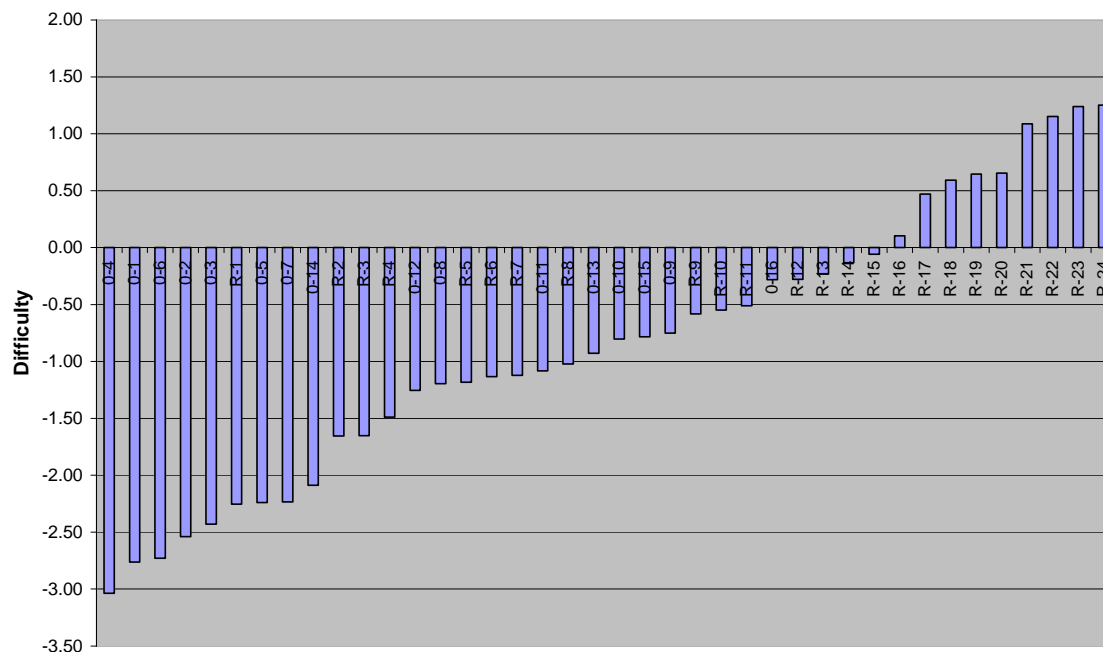
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

Figure 3. **Item difficulties for language and literacy items in the routing and level 1 child assessment forms**

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.
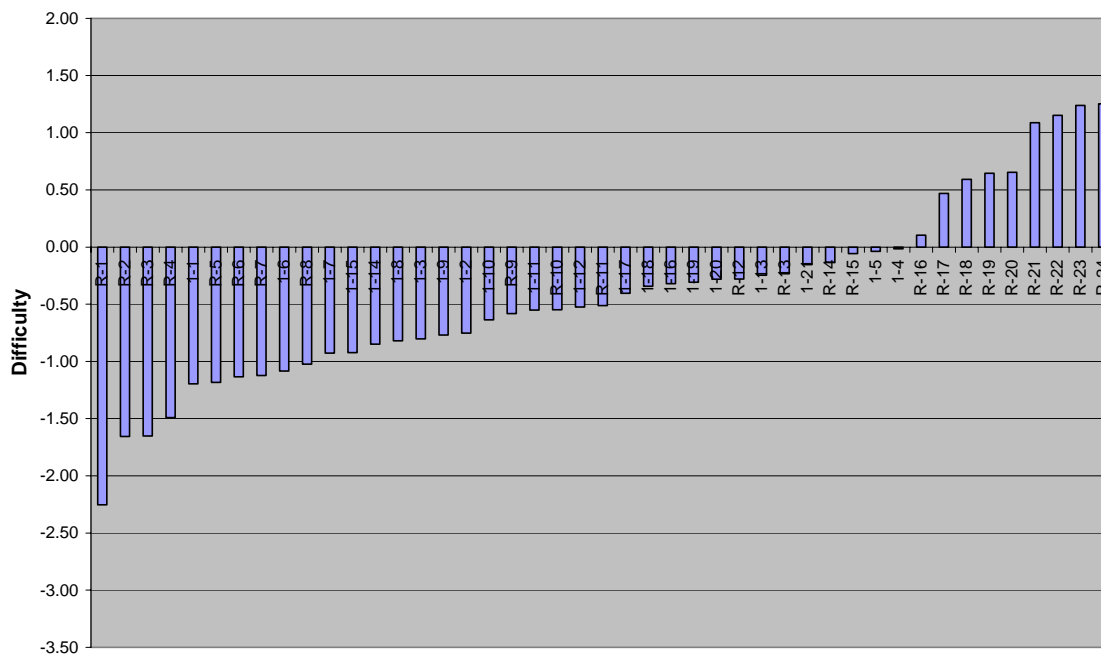
Figure 4. **Item difficulties for language and literacy items in the routing and level 2 child assessment forms**
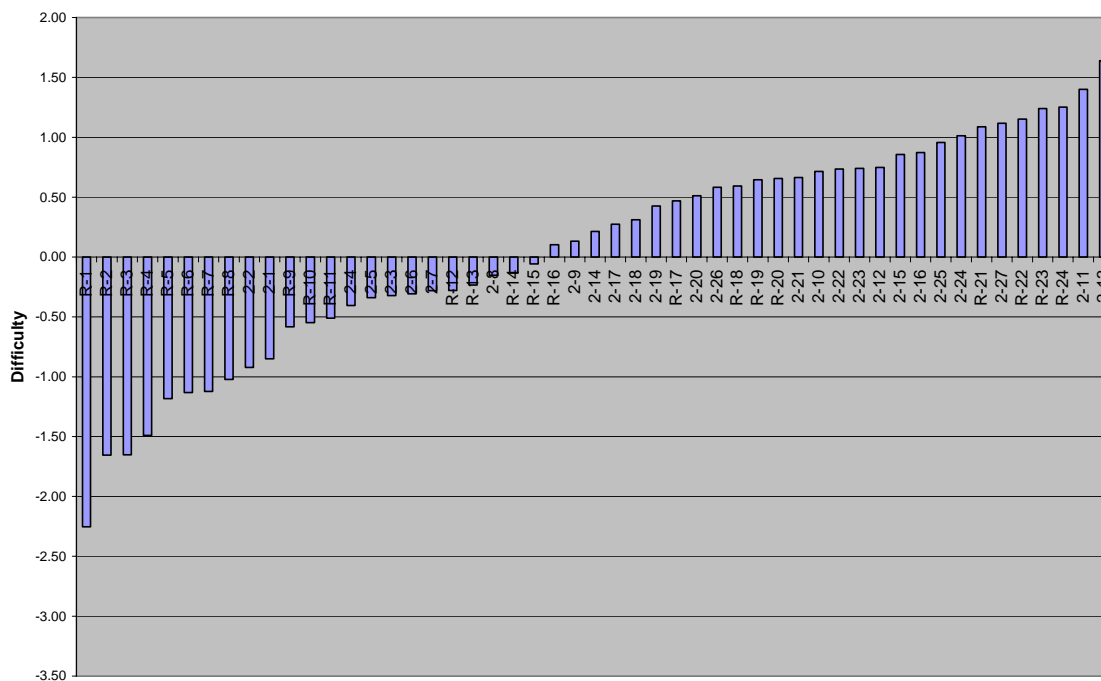

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

ECLS-B, K-1 Field Test Report Memo #4

An important component of the two-stage direct child assessment is that the items for the routing are selected to provide a common set of items for all children, regardless of their ability level, and to provide items that can identify child ability level, directing the assessment to the appropriate second-stage of the assessment. This is built into the assessment design by the development of cut-scores, which indicate which second stage form a child should completed based upon the number of items correct in the routing form. The cut scores and projected percentage of children directed to each second stage form for the language and literacy items is shown in Table 5.

Table 5.   Cut scores for, and projected percentage of children routed to, each second stage form, language and literacy

Second stage form	Routing Score that Directs to Second-Stage Form	% Routed to Second-Stage Form
Level 0	0-7	28
Level 1	8-13	52
Level 2	14-24	21

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

A second component of the assessment designed to focus items of specific difficulty to children based upon their skills, is the development of discontinue rules. Put simply, when a child is given multiple opportunities (i.e., items) to show a skill but does not show that skill, additional items of that type are not administered. Additionally, when there is a known sequence of skills, discontinue rules eliminate from the assessment for a child those items which would likely be beyond the child's skill and therefore would have a high probably of being incorrect.

The items selected for inclusion in the Kindergarten '06 and '07 data collection reflect consideration given to both the established framework for child assessment used across ECLS-K and ECLS-B cohorts, and the goal of maintaining overlap between the ECLS-B kindergarten instrument and the ECLS-B preschool instrument and ECLS-K kindergarten instrument. The framework for the original ECLS-K assessments was based on defining the required content categories as an estimated percentage of testing time, not as a count of items. (In fact, neither ECLS-B nor ECLS-K had access to actual individual item timings.) The child assessment for the ECSL-B kindergarten rounds of data collection includes items that tap specific constructs in close approximation to the representation of the construct (as a percentage of time) in the ECLS-K (see table 5). The specific constructs are listed below, and indicated in the first column of table 5.

Table 5.   ECLS-K K-1 framework targets and items by content area

| | ECLS-K | | | ECLS-B | |
Category	Actual number of items	Percentage of items from actual forms	Percentage of time from frameworks	Total number of items	Percentage of items in proposed pool
Total	72	100	100	74	100
Basic skills	40	55	40	53	71
Vocabulary	8	11	10	7	10
Initial understanding	20	28	10	10	13
Developing interpretation	2	3	25	2	3
Personal reflection and response	0	0	10	0	0
Critical stance	2	3	5	2	3

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

The next three columns refer to the ECLS-K assessments. The second and third columns show the number and percentage of items administered in the forms. The fourth column lists the estimated percentage of time from the framework document. In the ECLS-B design we targeted for percentages that fell between those in columns 3 and 4. The last two columns list the number and percentage of items in the ECLS-B forms. The percentage of ECLS-B items (column 6) falls between the values in columns 3 and 4 in all content areas except for Basic Skills. Since the ECLS-B kindergarten assessment does not need to cover the range of ability levels through the spring of first grade, more of the lower difficulty items, such as those categorized as Basic Skills were added, while those more appropriate for first grade, such as those categorized as Initial Understanding were reduced. (All Initial Understanding items from the ECLS-K low and middle forms are included in the proposed ECLS-B kindergarten national forms.) It is our suggestion to not add in more of the Initial Understanding items, as they are more appropriate for higher ability levels, and to not remove any Basic Skills items, as they are more appropriate for the fall kindergarten ability range as well as for children not yet in kindergarten in 2006. In addition, the total number of items is higher in the ECLS-B than in the ECLS-K, due to the adding some easier, quicker items to the assessment.

The ECLS-B preschool national items included in the proposed kindergarten forms, with the exception of the vocabulary items, were categorized as Basic Skills (per a phone conference regarding frameworks with NCES, RTI, and ETS). However, in the preschool national, these items were "sub-categorized" into more refined content areas. Table 6 lists the Basic Skills sub-categories, the targeted and actual numbers and percentages of items in the ECLS-B preschool national forms, and the numbers and percentages of items in the proposed ECLS-B kindergarten forms. (Although "Vocabulary" is not a Basic Skills subcategory, it is a content area from the preschool frameworks and is therefore included in table 6 for completeness.)

Table 6.   ECLS-B preschool framework targets and items by content area

| BS Subcategory | ECLS-B preschool national | | | | ECLS-B kindergarten proposed forms | |
| | Target | | Actual | | | |
	Number of items	Percentage of items	Number of items	Percentage of items	Number of items	Percentage of items
Total	81	100	73	100	60	100
English language skills/oral language	22	27	20	27	7	12
Phonological awareness	18	22	10	14	15	25
Letter and letter-sound knowledge	12	15	14	19	14	23
Print conventions	7	9	8	11	6	10
Word recognition	6	7	5	7	11	18
Vocabulary	16	20	16	22	7	12

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

The percentages in table 6 represent a reduced emphasis on language acquisition (English Language Skills/Oral Fluency/Vocabulary) and an increased emphasis on early literacy skills (letter and word, etc.) as children progress from the preschool year to kindergarten. The percentage of items by category in columns 3 and 5 (preschool target and actual) and column 7 (kindergarten) differs, but still shows coverage of the subcategories in the kindergarten assessment. Moving down column 7, fewer items from the English Language Skills/Oral Language subcategory are included in the kindergarten proposed forms. These include the Pre-LAS "Simon Says" and "Art Show" items, a portion of which were included in the kindergarten assessment. These items were very easy and it was not necessary to include all of them in the kindergarten forms. Onset rime items are included in the kindergarten assessment, thereby increasing the number of Phonological Awareness items. Similarly, the addition of ending sounds items increased the number of Letter and Letter-Sound Knowledge items. The number of Print Convention items remained about the same between preschool and kindergarten, while the number of Word Recognition items increased as several were added to the kindergarten forms. The number of vocabulary items is lower in order to match the ECLS-K K-1 framework in that category. Again, these are the preschool frameworks, and the purpose of this comparison with the kindergarten assessment is to demonstrate the variety of subcategories included, not to define what should be included in the kindergarten forms.

Developing a vertical scale that must span preschool to kindergarten and have optimal measurement properties throughout the achievement range calls for multiple test forms that vary in difficulty. These forms are tailored for individuals within a grade; however, the overall grade-level forms should reflect core curriculum elements for that particular grade. At the same time there must be overlapping items shared by forms within a year, as well as across years. These linking items tie the vertical scale together both across forms within a year and across years. At least 20 to 30 percent of the items should overlap between adjacent years.

Tables 7 and 8 show the overlap of items across forms within the proposed kindergarten assessment and with the ECLS-B preschool national, kindergarten field test, and ECLS-K K-1 assessments. The 24 routing items, taken by all children, also serve as common items for linking across forms. (Only one item [read aloud: mom] is shared across all three forms, as the routing items serve the same purpose.) Based on the rule of thumb listed above, there are ample numbers of items across forms and across assessments to create both horizontal (i.e., within-grade, and cross-cohort) and vertical (i.e., longitudinal) scales.

Table 7. Number of items overlapping across forms

Form	Number of unique overlapping items
Levels 0 and 1	5
Levels 1 and 2	7
Levels 0, 1, and 2	1

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

Table 8. Number of items on the ECLS-B kindergarten assessment overlapping with other assessments

Assessment	Number of unique overlapping items
ECLS-B Preschool National	35
ECLS-B K Field Test	40
ECLS-K K-1 National	50
All of the above	14

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

3.   Mathematics Items

3.1 Completion Rates

All children ($N = 303$) completed the mathematics items on the K-1 field test. As with the language and literacy assessment, it should be noted that the assessment was designed to present children with items based upon their performance on the assessment. In practice, this resulted in children completing items that estimated their skill level, while those items that were either below or above their performance level were presented only until the child's performance level was identified. Therefore, many children skipped items that were "too easy" and did not complete items that were "too hard" beyond established criteria within the assessment. Such rules result in more efficiency during assessment (i.e., time is not taken in delivering items the child would be expected to pass, based upon his or her performance) and in keeping the child engaged by ending the assessment once the items became more difficult, but before the child would likely become distressed by an inability to complete these harder items.

3.2 Completion Time

The median completion time for the language and literacy assessment was 14.02 minutes. The mean and median time of the child math assessment, as well as the inter-quartile range, are given in table 9. Many children were routed out of the harder items, lowering the average time of assessment.

Table 9. Child assessment length

Child assessment cases—mathematics			
N	Mean time (in minutes)	Median time (in minutes)	Inter-quartile range
303	15.00	14.02	4.88

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

3.3 Item Analysis

We began by using IRT ability estimates to define target difficulty ranges for different forms of the test. The ability (theta) estimates for the ECLS-K fall kindergarten sample, from the pooled analysis above, is used as a best estimate of the ECLS-B fall kindergarten sample, with a mean = 0.035, and standard deviation = 0.796. To the extent that the pooled ECLS-K fall kindergarten data is representative, three standard deviations above and below the mean (about -2.36 to +2.43) should include about 99 percent of the ECLS-B kindergarten sample. This range also covers the (pooled) preschool ability estimate to 1.5 standard deviations (0.837) below its mean (-1.10), which provides coverage to those children in the lower ability range, possibly such as those not entering kindergarten in fall 2006.

The range of -2.36 to +2.43 defines not only the ability range of the children, but also the corresponding difficulty (b) parameters of the items required for the assessment. Of course, the parameter estimates are not precise or permanent, and will change to some extent in the national administration. There are many reasons for possible changes in the parameter estimates: the assortment of items and the order in which they will be given, the number and location of practice items, discontinue rules, etc. As a precaution against encountering floor and ceiling effects in the national assessment (sets of items that are *all* too hard or too easy for some of the children taking them) we recommend extending the difficulty range of the items at both the low and high end. We have included some items with (b) parameters below -2.36 on the proposed low second stage form, and some above +2.43 on the high form.

We reviewed the psychometric characteristics of the items and deleted any items that were unsatisfactory with respect to the quality criteria described above. Then we sorted the items according to category and presentation. For example, "Number sense, properties and operations" items were presented in several different formats. We reviewed the difficulty statistics for the items within each content/presentation type, and classified each set as suitable for the routing test, the low, middle or high form, according to whether the majority of the items in each set had an appropriate spread of difficulty. We compared different presentations of the same content, and

where there was redundancy, selected the item sets with the strongest characteristics. In general we ordered the items so that content was presented in increasing order of average difficulty (although most content areas had a spread of difficulty within them) and we grouped item types.

The form design was based on what was used in the ECLS-K K-1 assessments, with several modifications. The ECLS-B kindergarten assessment does not require the very difficult items present on the high form of ECLS-K, which were needed to assess children in the spring of first grade. The ECLS-K high form was therefore removed, and what were the ECLS-K middle and low forms became the ECLS-B Level 2 and Level 1 forms, respectively. Again, several modifications were made, such as including some of the harder items from the ECLS-K high form, rearranging the items, removing some, etc. A Level 0 form was assembled, which includes several of the routing and high form items from the ECLS-B preschool national assessment, in addition to some ECLS-K low form items. The routing test was also modified to include all items from the ECLS-K proficiency levels 1 to 4, with those from level 5 added to the ECLS-B Level 2 form.

The distributions of thetas described above define the range of ability to be targeted by the test forms. The IRT difficulty parameters for the pool of available items are calibrated on the same scale as the abilities. Thus the process of choosing test items relies on matching the difficulty of the items to ability of the test takers. To optimize the measurement accuracy of the tests, the selected items should be approximately equally spaced along the ability/difficulty scale. Table 10 below shows the peak difficulty ranges for the set of items to be administered to each child, that is, the routing test plus one second stage form (50 percent of the abilities expected to fall in the Routing + Level 1 category, 25 percent lower, and 25 percent higher), the number of items in the peak range, the full range of difficulty per form, and the total number of items per form. As indicated above, to avoid floor and ceiling effects, the abilities in each of the forms extend beyond the peak difficulty ranges. (The items outside the peak range are a result of including the full range of routing items, the intentional addition of items to extend difficulties beyond the peak range, and the addition of items to provide overlap between forms needed to support development of a common score scale.)

Table 10. Peak and full difficulty ranges, routing + second stage

	Routing + Level 0	Routing + Level 1	Routing + Level 2
Total number of items	33	37	42
Peak difficulty range	-2.36 to -0.50	-0.50 to +0.57	+0.57 to +2.43
Number of items in peak range	17	13	21
	5 Routing	5 Routing	5 Routing
	12 Level 0	8 Level 1	16 Level 2
Full range of difficulty	-2.83 to +1.77	-2.33 to +1.77	-2.20 to +2.83

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

3.4   Recommendation for the National Assessment

The child mathematics assessment in Kindergarten '06 and '07 rounds will again use a two-stage design, with all children being administered the routing form, which will direct them to the appropriate second-stage form based upon their ability level. The recommended math items include 17 routing items, as well as 16 items at level 0, 20 items at level 1, and 25 items at level 2. These items are drawn based upon their psychometric strength, as revealed through IRT analyses. The mathematics items for each form of the assessment, with parameters for discrimination, difficulty, and guessing, are shown in table 11.

Table 11. Item run parameters for mathematics items to be included in ECLS-B Kindergarten '06 and '07 child assessment, by routing form and item order

| Proposed form	Short name	Item description	Combined run parameters		
			Discrimination	Difficulty	Guessing
ROUTING					
R-1	SQUARE	which of these shapes is a square?	0.617	-2.204	0.119
R-2	SM-LG-SM	Match pattern: small-large-small	0.943	0.191	0.324
R-3	3BANANAS	point to picture of 3 bananas	0.657	-1.871	0.094
R-4	6BANANAS	how many bananas in all (6)	0.739	-0.098	0.001
R-5	COUNT10	Count 20 circles: rescore--correct if 10+	0.751	-1.821	0.001
R-5	COUNT 20	Count 20 circles	0.915	-0.098	0.001
R-6	NUMBER 4	What is this? (number 4)	1.881	-1.261	0.001
R-7	NUMBER 9	What is this? (number 9)	1.677	-0.408	0.001
R-8	NUMBER23	What is this? (number 23)	1.373	0.491	0.001
R-9	STICKBAT	Which stick is shorter than bat	0.645	-0.821	0.060
R-10	3RD LINE	Point to the person 3rd in line	1.226	0.519	0.001
R-11	_ 78910	What goes in blank: __ 7 8 9 10	1.286	0.576	0.001
R-12	51015_25	What goes in blank: 5 10 15 __ 25	1.423	1.765	0.001
R-13	3+2 CARS	3 cars plus 2 more, how many	0.918	0.548	0.001
R-14	5-1ORANG	5 oranges, ate 1, how many left	1.231	1.292	0.122
R-15	2+5MARBL	5+2 marbles in bag	0.997	1.263	0.001
R-16	3+7PENNY	had 3 pennies, dad gave 7 more	1.316	1.587	0.001
LEVEL 0					
0-1	MOREBALL	Which has more balls	0.644	-2.834	0.001
0-2	FEWBALLS	Which has fewer baseballs	0.536	-1.402	0.001
0-3	2CRAYONS	how many crayons on page (2)	1.353	-2.326	0.001
0-4	STARS5	Count 5 stars	1.564	-1.662	0.001
0-5	STARS12	Count 12 stars	1.441	-1.022	0.001
0-6	HOUSE	Point to a triangle in this house	0.830	-1.802	0.001
0-7	FINGERS4	Show me 4 fingers	1.075	-1.583	0.001
0-8	LASTBIKE	Point to the last bicycle	0.616	-1.266	0.150
0-9	TALLTREE	Point to the shortest tree	0.555	-1.036	0.205
0-10	NUMBER 7	What is this? (number 7)	1.733	-0.852	0.001
0-11	NUMBER17	What is this? (number 17)	1.369	0.250	0.001
0-12	FOURDUCK	Which picture matches 4 ducks	0.873	-0.768	0.305
0-13	#CHOC	graph: how many chocolate	0.915	-0.418	0.001
0-14	#VANILLA	graph: how many vanilla	0.867	-0.692	0.001
0-15	#STRAW	graph: how many strawberry	0.756	-1.071	0.001
0-16	PNTBRUSH	how many inches is paintbrush	1.092	-0.275	0.210
LEVEL 1					
1-1	2CRAYONS	how many crayons on page (2)	1.353	-2.326	0.001
1-2	STARS5	Count 5 stars	1.564	-1.662	0.001
1-3	STARS12	Count 12 stars	1.441	-1.022	0.001
1-4	HOUSE	Point to a triangle in this house	0.830	-1.802	0.001
1-5	NUMBER 7	What is this? (number 7)	1.733	-0.852	0.001
1-6	NUMBER17	What is this? (number 17)	1.369	0.250	0.001
1-7	#CHOC	graph: how many chocolate	0.915	-0.418	0.001

See notes at end of table.

Table 11. Item run parameters for mathematics items to be included in ECLS-B Kindergarten '06 and '07 child assessment, by routing form and item order—Continued

Proposed form	Short name	Item description	Combined run parameters		
			Discrimination	Difficulty	Guessing
1-8	#VANILLA	graph: how many vanilla	0.867	-0.692	0.001
1-9	PNTBRUSH	how many inches is paintbrush	1.092	-0.275	0.210
1-10	LG-SM-SM	match pattern: large-small-small	1.096	0.124	0.296
1-11	4LINES	match pattern: 4 lines = square	0.661	0.323	0.171
1-12	000X	match pattern: same/same/same/different	0.795	0.386	0.205
1-13	HALFOVAL	match pattern: half oval/half triangle	0.787	0.771	0.286
1-14	DANNY	5-2 bananas	0.962	0.089	0.001
1-15	CATS	3+1 cats	1.214	0.390	0.178
1-16	EBONY	3-1 books	0.975	0.570	0.198
1-17	2+3STICK	has 2 sticks, found 3 more; # total?	1.018	0.653	0.001
1-18	RACHEL	2+2 butterflies	1.289	0.711	0.255
1-19	8-6CRAYN	8 crayons, take away 6; # left?	0.861	1.120	0.001
1-20	#BUGS	estimate how many bugs	0.992	0.712	0.213
LEVEL 2					
2-1	#CHOC	graph: how many chocolate	0.915	-0.418	0.001
2-2	#VANILLA	graph: how many vanilla	0.867	-0.692	0.001
2-3	PNTBRUSH	how many inches is paintbrush	1.092	-0.275	0.210
2-4	LG-SM-SM	match pattern: large-small-small	1.096	0.124	0.296
2-5	4LINES	match pattern: 4 lines = square	0.661	0.323	0.171
2-6	000X	match pattern: same/same/same/different	0.795	0.386	0.205
2-7	HALFOVAL	match pattern: half oval/half triangle	0.787	0.771	0.286
2-8	SHAPES	match pattern: large-small-small-large	0.510	1.439	0.216
2-9	PATTERN	extend pattern of shapes	0.879	1.374	0.181
2-10	EBONY	3-1 books	0.975	0.570	0.198
2-11	2+3STICK	has 2 sticks, found 3 more; # total?	1.018	0.653	0.001
2-12	8-6CRAYN	8 crayons, take away 6; # left?	0.861	1.120	0.001
2-13	2+5CIRCL	2 circles, drew 5 more; # total?	1.068	1.462	0.001
2-14	#BUGS	estimate how many bugs	0.992	0.712	0.213
2-15	12 BY 2S	count to 12 by two's	1.311	1.494	0.001
2-16	COST$10	Lunch $2 x 5 children	1.375	2.442	0.001
2-17	15/5CARS	15 children, 5 per car	1.390	2.565	0.001
2-18	8/2CANDY	has 8 cents; candy costs 2 cents	1.281	2.834	0.001
2-19	13_79	What goes in blank: 1 3 __ 7 9	1.300	2.314	0.001
2-20	2 + 2	what number goes in box: 2+2	2.216	0.774	0.001
2-21	1 + 7	what number goes in box: 1+7	0.927	0.902	0.001
2-22	3 + 4	what number goes in box: 3+4	1.373	1.248	0.001
2-23	7 - 3	what number goes in box: 7 - 3	1.738	1.835	0.001
2-24	12 + 6	what number goes in box: 12+6	1.167	2.164	0.001
2-25	17 - 4	what number goes in box: 17-4	1.615	2.429	0.001

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

Because each child will receive items in the routing form in addition to the appropriate second-stage form, we would expect item difficulty across combinations of forms (routing plus level 0, routing plus level 1, routing plus level 2) to demonstrate greater difficulty in the higher forms than the lower forms. Item difficulty for routing items plus items in form 0 (see Figure 5), routing items plus level 1 (see Figure 6), and routing items plus level 2 (see Figure 7) are shown below.

Figure 5.   Item difficulties for mathematics items in the routing and level 0 child assessment forms.
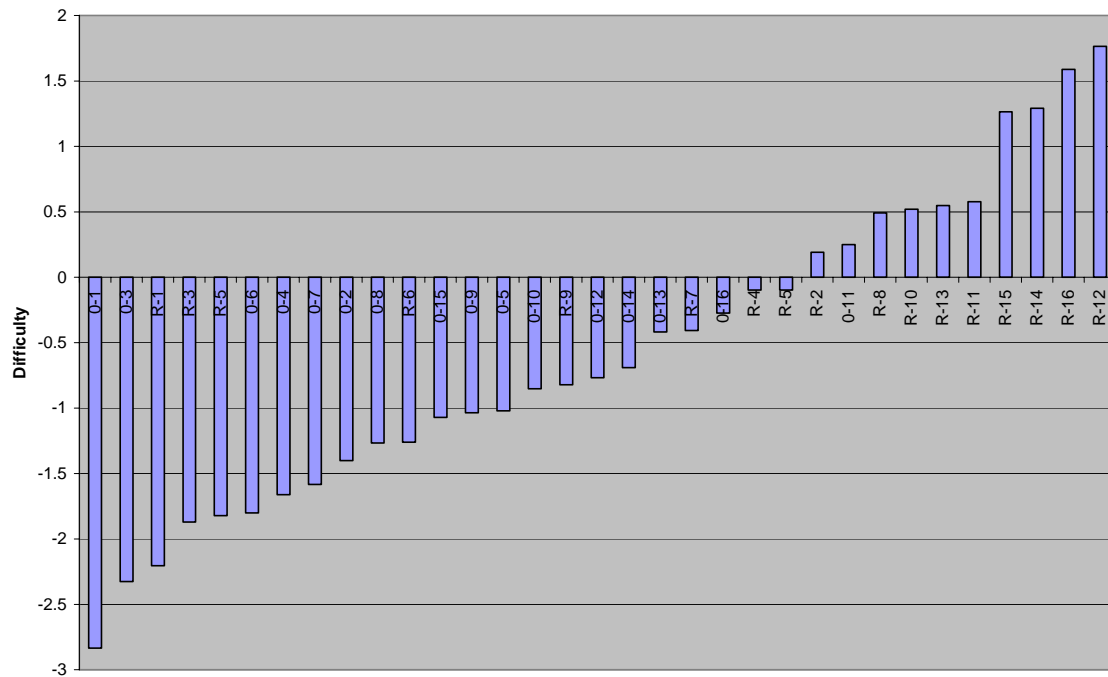
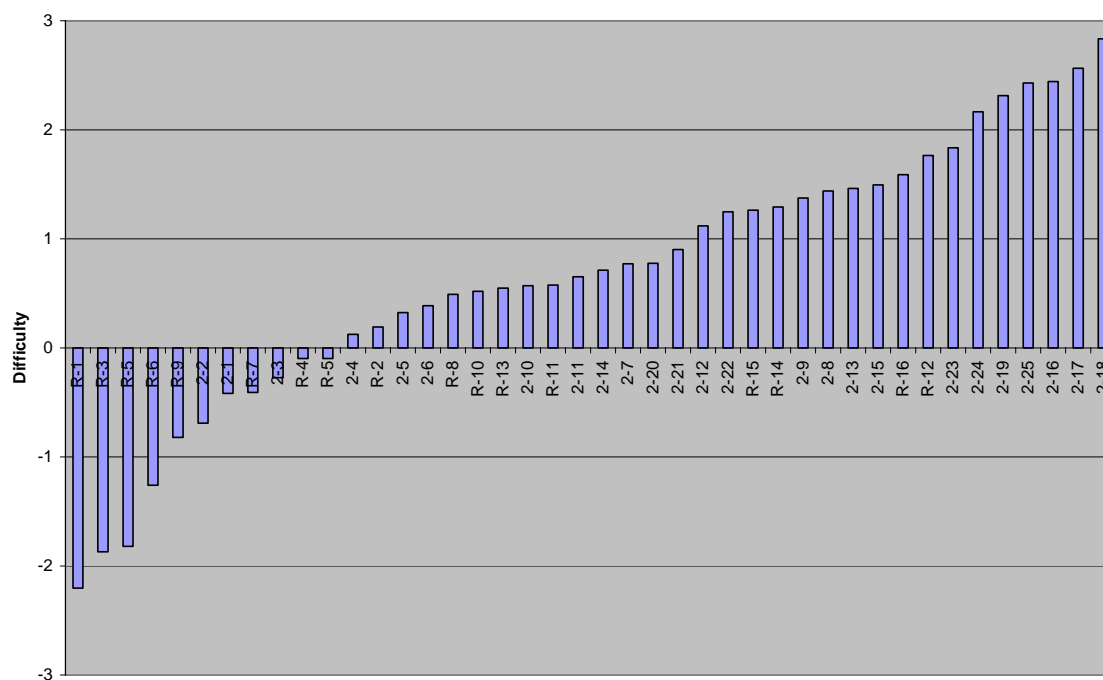**Figure 6. Item difficulties for mathematics items in the routing and level 1 child assessment forms.**
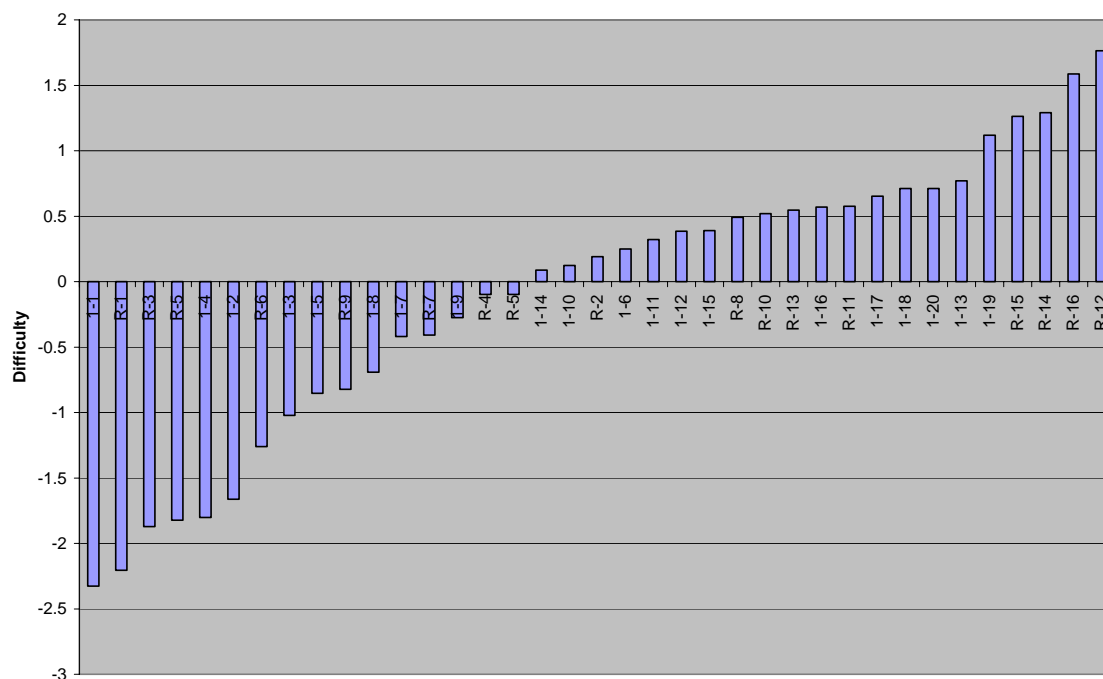
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

Figure 7. Item difficulties for mathematics items in the routing and level 2 child assessment forms.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

These items were selected with consideration given to both the established framework for child assessment and the goal of maintaining overlap between the ECLS-B K-1 instrument and the ECLS-B preschool instrument and ECLS-K kindergarten instrument. The framework for the original ECLS-K assessments was based on defining the required content categories as an estimated percentage of testing time, not as a count of items. (In fact, neither ECLS-B nor ECLS-K had access to actual individual item timings.) The child assessment for the ECSL-B kindergarten rounds of data collection includes items that tap specific constructs in close approximation to the representation of the construct (as a percentage of time) in the ECLS-K (See table12). The specific constructs are listed below, and are indicated in the first column on table 12.

- NSPO, Number sense, properties and operations
- MEAS, Measurement
- GSS, Geometry and spatial sense
- DSP, Data analysis, statistics and probability
- PAT, Patterns, algebra and functions

Table 12. Framework targets and items by content area

	ECLS-K			ECLS-B	
	Actual number of items	Percentage of items from actual forms	Percentage of time from frameworks	Total number of items	Percentage of items in proposed pool
Total	65	100	100	58	100
NSPO	49	75	50	41	71
MEAS	3	5	15	3	5
GSS	2	3	5	4	7
DSP	5	8	10	3	5
PAF	6	9	20	7	12

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

The next three columns refer to the ECLS-K assessments. The second and third columns show the number and percentage of items administered in the forms. The fourth column lists the estimated percentage of time from the framework document. In the ECLS-B design we targeted for percentages that fell between those in columns 3 and 4. The last two columns list the number and percentage of items in the ECLS-B forms. The percentage of ECLS-B items (column 6) falls between the values in columns 3 and 4 in all content areas except for DSP and GSS. Those items on the low and middle forms of ECLS-K categorized as DSP were included in the ECLS-B assessment. The others existed only on the ECLS-K high form, and were therefore not included. Similarly, more GSS items existed on the lower forms of the ECLS-K and on the ECLS-B. It is our suggestion to not add in more of the DSP items, as they were more appropriate for higher ability levels, and to not remove any GSS items, as they are more acceptable for the fall kindergarten ability range.

The reduction in total number of items administered between the ECLS-B and ECLS-K assessments is a result of elimination of the ECLS-K high form (which was taken mostly by ECLS-K first-graders) from the ECLS-B assessments. This form in ECLS-K not only had many very difficult items, but was also quite long. Eliminating this form results in a total reduction in the number of items administered. We believe that this number of items is sufficient for maintaining accuracy of measurement while minimizing burden.

Developing a vertical scale that must span preschool to kindergarten and have optimal measurement properties throughout the achievement range calls for multiple test forms that vary in difficulty. The forms are tailored for individuals within a grade; however, the overall grade-level forms should reflect core curriculum elements for that particular grade. At the same time there must be overlapping items shared by forms within a grade, as well as across grades. These linking items tie the vertical scale together both across forms within a grade and across grades. At least 20 to 30 percent of the items should overlap between adjacent grades.

Tables 13 and 14 show the overlap of items across forms within the proposed kindergarten assessment and with the preschool national and ECLS-K K-1 assessments. The 16 routing items, taken by all children, also serve as common items for linking across forms. Based on the rule of thumb listed above, there are ample numbers of items across forms and across assessments to create both horizontal (i.e., within-grade, and cross-cohort) and vertical (i.e., longitudinal) scales.

Table 13.   Number of items overlapping across forms

Form	Number of unique overlapping items
Levels 0 and 1	6
Levels 1 and 2	3
Levels 0, 1, and 2	8

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

Table 14.   Number of items on the ECLS-B kindergarten assessment overlapping with other assessments

Assessment	Number of unique overlapping items
ECLS-B Preschool National	31
ECLS-B K Field Test	35
ECLS-K K-1 National	45
All of the above	18

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), Combined Kindergarten-1st Grade Field Test.

4.   Physical and Motor Skill Items

The field test included neither the physical measurement items nor the fine motor skills items included in the ECLS-B preschool national study. However, both physical measurements and fine motor skills will be included in the kindergarten national data collection. The child assessment administered in the field test included a subset of the gross motor items from the ECLS-B preschool national study. The gross motor items included standing jump, skip, and catch a bean bag.

The gross motor skills items were included in the field test to break up the cognitive assessment, as well as to gain important information on the physical development of the child. These items performed as expected and are recommended for retention in the child assessment. Based upon data from the ECLS-B preschool national study and the fact that children in the kindergarten national study will be older, we recommend three minor modifications in assessment procedures. The first is to instruct field interviewers to take measures of child middle upper-arm circumference and head circumference (if applicable) while the child is standing, rather than seated on the mother's lap, unless the child is unable to stand. It is our judgment that this procedure will be more comfortable for the child and field interviewer. The second modification is in the range of expected values for each of the physical measures to reflect expectations for each measurement derived from National Center for Health Statistics growth charts. The final modification is an additional instruction that measures taken twice should be approximately the same. Previously field interviewers were coached that these measures should be within 5 percent, but we think data accuracy will be improved by providing field interviewers with a specific value by which these two measures should not differ.

Fine motor items were not included in the ECLS-B K-1 field test; however, data on these items were available from the ECLS-B preschool national study. Based upon a review of these data, many items appeared to indicate ceiling effects. More than 93 percent of children in the national preschool data collection passed 9 of the 11 items. On one additional item (build a tower, FM01) nearly two-thirds passed (73.5 percent). On only one item (build a gate, FM02) was there little evidence of floor or ceiling effects (approximately 45 percent of children passed). Based upon this, we recommend extensive modification of the fine motor items.

We propose to modify this part of the assessment to more closely match the assessment of these skills used in the ECLS-K data collection. Specifically, the fall ECLS-K K fine motor items included the square, triangle, open-square with circle on the right lower corner, the asterisk, the block gate using five cubes, and the draw-a-person items. The use of these items will allow a direct link between ECLS-B and ECLS-K cohorts in kindergarten, and the inclusion of several items from the ECLS-B Preschool National study (i.e., build a gate, square, triangle, asterisk) will provide for linkage between ECLS-B preschool and kindergarten data collection.